# BUFFER INSTALLATION METHOD OF SUPPORTING A DETECTION-BASED AND AVOIDANCE-BASED CONSISTENCY MAINTENANCE POLICY IN A SHARED DISK-BASED MULTI-DBMS

## BACKGROUND OF THE INVENTION

### Field of the Invention

**[0001]** The present invention relates to a cache coherency maintaining method for a database management system (DBMS) operating in a shared disk multisystem, and more particularly, to a buffer allocation method of supporting a detection-based consistency maintenance policy and an avoidance-based consistency maintenance policy in a shared disk-based multi-DBMS improved in which an optimal procedure for each type is selectable by using the characteristic of the DBMS controlled coherently by units of table, block and record and the respective procedures coexist to obtain more improved performance.

### Discussion of the Related Art

**[0002]** Generally, when it is required to cache the same data in two or more systems, a cache consistency maintenance procedure is used to prevent cached contents from being inconsistent with each other, and is used as an essential procedure in a shared disk-based file system, a client server DBMS, a parallel DBMS, a cluster DBMS, etc.

**[0003]** A database management system 10 managed in the conventional single system includes a DBMS 11, a file manager 12, a buffer manager 13, a buffer 14, a storage device 15, a log device 16 and a locking manager 17. In this system, since a unit of cache is a block used as a unit for input/output from/to a storage device, the cache consistency maintenance is performed in the block unit. In other words, when an access to a buffer is

performed to read and update a disk block in the specific system, the buffer manager is ensured to provide the recent contents.

**[0004]** Various methods for this scheme are being proposed currently. Rahm E. divided a block-based consistency maintenance scheme into a detection-based scheme and an avoidance-based scheme in the paper entitled "Concurrency and Coherency Control in Database Sharing Systems", Technical Report ZRI 3/91, University of Kaiserslautern, Dept. of Computer Science.

**[0005]** In the detection-based scheme, upon access to a buffer, if there is a block cached in the buffer, it is checked whether or not the block is proper to use according to some criterion. If the block is improper to use, the block of a most recent version is loaded by a predetermined procedure. In this scheme, when updated, only the information that the block has been updated is propagated to each system or kept at a shared area to keep the older version of the block cache. So, update propagation does not cost much.

**[0006]** There is a method to track whether or not a block was updated. In this method, the number of a recent version is kept at a specific location and is transferred to the buffer manager when accessing the buffer. The transferred number is compared with the number of a version of the cached block.

**[0007]** The paper discloses that the performance can be improved when the detection-based scheme is combined with locking by record. Generally, there are a few records in a block. Even though a specific record of a specific block is being updated, the block of the old version cached in the system can be used if another system accesses a record of the same block that is not being updated.

**[0008]** In this case, as a result of a record locking, the version number of the block required to access the record is transferred and this information is used to check validity of the block cached when buffer is loaded. This procedure is disclosed in the USP 5,327,556 entitled "Fast intersystem page transfer in a data sharing environment with record locking".

In the patent, the L-lock for record locking and the P-lock for authorizing to load page on buffer are used to describe the record level detection-based scheme.

**[0009]** However, this detection-based scheme cannot obtain information on version from lock result when accessing a record accompanied with table-based lock. In other words, table-based lock is obtained once before accessing all of the records in the table. Accordingly, whenever accessing a record, the version of the corresponding block should be obtained which is very inefficient.

**[0010]** In the avoidance-based scheme, when updating a buffer, caches of all the systems storing those of old version are invalidated and updated. The contents of the caches are propagated to the systems. This scheme costs much to update but it ensures that the blocks loaded in the buffers are always recent versions.

**[0011]** According to this scheme, if a specific block is being updated in a specific system, accessing the block is delayed in all the other systems until the block is updated completely. Instead, since it is not required to check the validity of the cached block, the procedure of obtaining a version number is not required before a buffer is allocated.

**[0012]** The avoidance-based scheme is proper to control consistency of the information read and updated block by block in DBMS. This information includes a system catalog block, index block, etc. The avoidance-based scheme is proper since these blocks are not allowed to be accessed.

**[0013]** The avoidance-based scheme can be used as a consistency maintenance method when accessing the record while locking a table. When this scheme is employed, it is not required to obtain the version information of the recent block before loading a buffer.

**[0014]** In the general DBMS, multi-granularity locking is employed in which table-based locking and record-based locking are used simultaneously. Therefore, to manage the table, detection-based scheme and avoidance-based scheme are used simultaneously according to a unit of locking.

[0015] These two methods should be applied interworking with each other using the same buffer space. If these two methods do not interwork with each other and use separate buffer spaces, the predictable problem is as follows.

[0016] First, when the specific block is allocated in the avoidance-based scheme, the specific block may be allocated in the same system and the other system in the detection-based scheme. In this case, the consistency of the cache of the database is deteriorated

[0017] Second, when the specific block is allocated in the detection-based update mode, it can be allocated in the avoidance-based read mode. So, problems may occur in the avoidance-based consistency control.

[0018] Third, when the specific block is allocated in the avoidance-based update mode, other systems can access the specific block in the detection-based read mode. So, problems may occur in the avoidance-based consistency control.

[0019] Accordingly, in the procedure to maintain cache consistency for DBMS, the two schemes should be supported and interwork with each other but there is no consistency maintenance procedure to meet these requirements now.

## SUMMARY OF THE INVENTION

[0020] Accordingly, the present invention is directed to a buffer installation method supporting a detection-based consistency maintenance policy and avoidance-based consistency maintenance policy in a shared disk-based multi-DBMS that substantially obviates one or more problems due to limitations and disadvantages of the related art.

[0021] An object of the present invention is to provide a buffer installation method supporting a detection-based consistency maintenance policy and avoidance-based consistency maintenance policy in a shared disk-based multi-DBMS to make the detection-based consistency maintenance scheme and the avoidance-based consistency maintenance

scheme interwork with each other and select one of them to use according to necessity and make it possible to access records by multi-granularity locking to enhance its performance.

**[0022]** Additional advantages, objects, and features of the invention will be set forth in part in the description which follows and in part will become apparent to those having ordinary skill in the art upon examination of the following or may be learned from practice of the invention. The objectives and other advantages of the invention may be realized and attained by the structure particularly pointed out in the written description and claims hereof as well as the appended drawings.

**[0023]** To achieve these objects and other advantages and in accordance with the purpose of the invention, as embodied and broadly described herein, the present invention provides a buffer installation method to select a consistency maintenance scheme when installing a buffer, and designating a version number of the block to install when selecting a detection-based scheme.

**[0024]** In another aspect of the present invention, a buffer installation method supporting a detection-based and avoidance-based consistency maintenance policy in a shared disk-based multi-DBMS comprises the steps of: (a) when a page identifier, an access mode (read, write) and a consistency maintenance scheme (detection, avoidance) are selected and a buffer is requested to be allocated, calculating a buffer locking mode required based on a SMTBM matrix; and (b) requesting a global locking manager to lock a buffer in the calculated buffer locking mode in case that an the obtained buffer locking mode is less than the calculated buffer locking mode or a version of a loaded block is lower than a required version when detection-based consistency maintenance scheme is selected, and approving buffer allocation otherwise, wherein a detection-based consistency maintenance scheme and an avoidance-based consistency maintenance scheme are integrated in a single procedure to interwork with each other.

[0025] In another aspect of the present invention, a method of processing a global locking request in a DBMS operated in a shared disk-based multi-system comprises the steps of: (a) obtaining a locking by an update authority (WX, X) in a system that has obtained a requested locking, transferring a corresponding block to a system that cached the corresponding block, and requesting to update a lock authority; (b) determining whether the system is not compatible to a requested lock according to a BLCM matrix in the system that has obtained the requested lock in a read mode (WS, S); and (c) instructing a system to update the lock authority, the system being determined not to be compatible.

[0026] In another aspect of the present invention, a method of a global locking manager for processing a locking authority update request and a block transfer request in a DBMS operated in a shared disk-based multi-system comprises the steps of: (a) when a current system has obtained a locking by an update authority (WX, X) and a current block is updated, writing a log forcedly about the current block based on write ahead logging (WAL) and writing a corresponding block on a disk or transferring the corresponding block through a transfer path; (b) updating a currently owned buffer locking mode to satisfy a buffer locking mode requested by a remote system using a BLCM matrix; and (c) removing a corresponding block completely when returning a buffer locking as a result of the step (b), and completing to update an ownership otherwise.

[0027] It is to be understood that both the foregoing general description and the following detailed description of the present invention are exemplary and explanatory and are intended to provide further explanation of the invention as claimed.

## BRIEF DESCRIPTION OF THE DRAWINGS

[0028] The accompanying drawings, which are included to provide a further understanding of the invention and are incorporated in and constitute a part of this

application, illustrate embodiment(s) of the invention and together with the description serve to explain the principle of the invention. In the drawings:

**[0029]** Fig. 1 is a block diagram to illustrate the database management system operated in the conventional single system;

**[0030]** Fig. 2 is a block diagram to illustrate the database management system operated in the environment of the present invention;

**[0031]** Fig. 3 is a flowchart to illustrate a buffer allocation method including cache consistency maintenance policy provided by the present invention;

**[0032]** Fig. 4 is a detailed flowchart to illustrate a buffer locking approval procedure in global locking manager to which the present invention is applied; and

**[0033]** Fig. 5 is a flowchart illustrating a procedure to update ownership authority of block of the present invention.

## DETAILED DESCRIPTION OF THE INVENTION

**[0034]** Reference will now be made in detail to the preferred embodiments of the present invention, examples of which are illustrated in the accompanying drawings.

**[0035]** Fig. 2 is a block diagram to illustrate the database management system operated in the environment of the present invention.

**[0036]** The database management system of the present invention has a first system 100 and a second system 200. The systems 100 and 200 have a log device 120 and 220, respectively. The two systems share a shared storage device 400.

**[0037]** The systems 100 and 200 of the present invention include a DBMS 101 and 201, a file manager 102 and 202, a buffer manager 103 and 203 and a local locking manager 104 and 204, respectively. A block receiver 130 and a block transmitter 230 are connected to each other through a communication path 501. The local locking managers 104 and 204 of each node are connected to each other through a global locking manager 300 and a

communication path 500. The global locking managers 104 and 204 are connected to a block transmitter 230 of the second system 200.

**[0038]** The buffer 110 and 210 to which the present invention is applied includes a block identifier (pid) 111 and 211, a load status (status) 112 and 212, a buffer locking mode (blmode) 113 and 213, a load version (ver) 114 and 214 and load data (data) 115 and 215, respectively.

**[0039]** The DBMS of the first system 100 analyzes a provided inquiry and instructs the file manager 102 to access a series of databases. The file manager 102 obtains a proper locking according to the type of data and requests the buffer manager to allocate a necessary block on a disk to access actual data. The buffer manager 103 receives the block of the recent version or reads the corresponding block from the disk with an aid of each module of the global locking manager 300 and the second system 200 to load on the buffer and approve allocation. In this process, various information in a buffer are referred to or updated.

**[0040]** Fig. 3 is a flowchart to illustrate a buffer allocation method including cache consistency maintenance policy provided by the present invention;

**[0041]** Referring to Fig. 3, the buffer allocation request of the present invention includes identifiers of a block, access modes, consistency maintenance modes and version numbers of the block. The access modes include READ and WRITE. The consistency maintenance modes include DETECT and AVOID. As long as the consistency maintenance mode is DETECT, the version number of the block is designated.

**[0042]** If buffer allocation is requested, the buffer manager 103 finds or allocates the corresponding buffer entry (S101). The buffer locking mode for processing the current request is calculated using the scheme mode to buffer lock mode matrix (SMTBM) as following Table 1 (S102). If the buffer locking mode determined like this is enough

compared with the already obtained buffer locking mode, the buffer allocation request is approved immediately.

**[0043]** If the' obtained locking mode is not enough compared with the requested buffer locking mode (S103) or the consistency maintenance mode is DETECT and the version of the current loaded block does not satisfy the requested version (S104), the global locking manager is requested for the buffer locking (S105), waits for until approval. In this status, when succeeding to receive a block (S106), the buffer allocation is approved. When failing to receive a block, the block is read from a disk (S107).

**[0044]** In the step S104, if the buffer locking is approved, it is checked whether the requested block is received. If succeeding to receive, the buffer allocation is approved immediately. Otherwise, the corresponding block is read from the disk to load on the buffer and the buffer allocation is approved (108).

**[0045]** Table 1

| Policy \ Access | Read | Write |
|---|---|---|
| Detection | WS | WX |
| Avoidance | S | X |

**[0046]** Table 1 illustrates the matrix SMTBM used in Fig. 3. Referring to Fig. 3, SMTBM is the matrix to find a buffer locking mode through requested consistency maintenance modes and access modes. The necessary buffer locking mode is the crossing portion of the row of the consistency maintenance mode and the column of the access mode requested by the matrix SMTBM.

**[0047]** The buffer locking modes include four kinds of modes, e.g., weak shared (WS), weak exclusive (WX), shared (S) and exclusive (X). The modes of WS and WX are used as the shared mode and the exclusive mode in detection-based buffer locking mode. The modes of S and X are used as the shared mode and the exclusive mode in avoidance-based buffer locking mode.

**[0048]** Fig. 4 is a detailed flowchart to illustrate a buffer locking approval procedure in global locking manager to which the present invention is applied.

**[0049]** Referring to Fig. 4, the present invention finds a locking entry corresponding to the requested block identifier in the global locking table and obtains the exclusive usage authority (S201). Then, the locking entry is searched (S202). The request (S203) to update transfer and ownership for the page requested by the system that obtained the buffer locking as the exclusive mode (WX, X) is sent to the block transmitter 230 of the system.

**[0050]** Then, all the systems that is not compatible to the requested buffer locking among the systems that obtained the buffer locking in the shared modes (WS, S) is determined through buffer lock compatibility matrix (BLCM) as Table 2 (S204). The request (S205) to update ownership authority of the system is sent to the block transmitter 230 of the system. If the process is completed, the buffer locking requested finally is registered on the locking table and the locking is approved (S206).

**[0051]** Table 2

| Request / Convention | WS | S | WX | X |
|---|---|---|---|---|
| NL | T | T | T | T |
| WS | T | T | T | F |
| S | T | T | F | F |
| WX | T | F | F | F |
| X | F | F | F | F |

**[0052]** Table 2 illustrates the matrix BLCM used in Fig. 4. Referring to Fig. 4, BLCM is the matrix to determine whether the buffer locking belonging currently to a specific system is compatible to the buffer locking requested currently. In the matrix BLCM, if the crossing portion of the row corresponding to the buffer locking belonging currently to a specific system and the column of the buffer locking requested currently is T,

it means that the ownership authority of the system does not have to update to approve the requested locking. If it is F, it means that the ownership authority of the system should be updated.

**[0053]** Fig. 5 is a detailed flowchart illustrating a procedure to transmit a block of the block transmitter of the second system to which the present invention is applied and to update ownership authority of block of the present invention. The block transmitter uses an identifier and a requested buffer locking mode of the block to be processed.

**[0054]** Referring to fig. 5, the identifier of the block is used to find a buffer entry (S301) to obtain the exclusive usage authority. If the loaded authority is the exclusive mode (WX, X) (S302), a log is written on a log device by the WAL protocol for the corresponding block and the corresponding block is transferred through a disk and a communication path (S303).

**[0055]** In step (S302), if the currently loaded authority is not the exclusive mode (WX, X), the buffer locking mode to be owned is determined using a buffer lock revocation matrix (BLRM) as following Table 3 (S304). If the determined locking mode is not No Lock (NL), it is completed immediately. Otherwise, the corresponding buffer entry is removed from the buffer completely and it is completed (S305 to S307).

**[0056]** Table 3

| Convention \ Request | WS | S | WX | X |
|---|---|---|---|---|
| WS | WS | WS | WS | NL |
| S | S | S | WS | NL |
| WX | WX | S | WS | NL |
| X | WX | S | WS | NL |

**[0057]** Table 3 illustrates the matrix BLRM used in Fig. 5. Referring to Fig. 5, BLRM is the matrix to determine the mode to update for the currently owned buffer locking to approve the buffer locking requested by other system. The buffer locking mode is the

crossing portion of the row corresponding to the currently owned buffer locking and the column corresponding to the requested buffer locking in the matrix BLRM. If this locking is NL, it means that the locking cannot be owned any longer and the corresponding buffer entry should be removed from the buffer completely.

**[0058]** As described above, the present invention allows to select and use one of the two kinds of the cache consistency maintenance schemes in the DBMS operated in the shared disk-based multi-system and to make these two kinds of the schemes interwork with each other. The following effects are expected.

**[0059]** First, the optimal cache consistency maintenance policy is employed for each granule corresponding to the essential multi-granularity locking in DBMS.

**[0060]** Second, the optimal cache consistency maintenance policy is employed for each case so that unnecessary block change between systems is reduced to improve the performance of the entire system.

**[0061]** Third, the two kinds of the cache consistency maintenance policy is integrated into the single buffer loading process to interwork so that the configuration of the system is simplified and implementation is easy.

**[0062]** Above-mentioned description is merely an example to illustrate a buffer allocation method supporting a detection-based and avoidance-based consistency maintenance policy in a shared disk-based multi-DBMS. The present invention is not bounded to the embodiments. It will be apparent to those skilled in the art that various modifications and variations can be made in the present invention. Thus, it is intended that the present invention covers the modifications and variations of this invention provided they come within the scope of the appended claims and their equivalents.